

■ LINKED (OPEN) DATA – BIBLIOGRAPHISCHE DATEN IM SEMANTIC WEB. BERICHT DER AG LINKED DATA AN DIE VERBUNDVOLLVERSAMMLUNG (16. MAI 2013)

von Patrick Danowski, Doron Goldfarb, Verena Schaffner, Wolfram Seidler

Inhalt

Einleitung

1. Was ist Linked Data?
2. Linked Data (LD) und Linked Open Data (LOD)
3. Die Open Data Strategien der Nationalbibliotheken
4. Projekte
5. Bibliothekarische Standards und Gremien
6. Normdaten und Culturegraph
7. Firmenpolitik
8. Implementierungsszenarien & Anwendungsfälle in Österreich
9. Bewertung/Fazit
10. Empfehlungen

Zusammenfassung: Linked Data steht für eine bestimmte Form der Veröffentlichung von Daten via Internet. Die zu Grunde liegende Idee ist es, Daten verschiedenster Provenienz, die derzeit teilweise gar nicht oder nur schwer zugänglich sind, in möglichst einheitlicher Form miteinander zu verknüpfen und dadurch in ihrer Gesamtheit abfragbar zu machen.

Dieser Bericht fasst die Entwicklungen im europäischen Raum, sowie strategische und technische Überlegungen der AG Linked Data hinsichtlich der Veröffentlichung von bibliothekarischen Daten des Österreichischen Bibliothekenverbundes (OBV) zusammen und schließt mit der gemeinsamen Übereinkunft, dass die Umsetzung von Linked Data-Prinzipien im OBV nur in Zusammenhang mit einer Diskussion über die damit einhergehende Veröffentlichung der Daten unter einer freien Lizenz angedacht werden sollte.

Schlagwörter: Linked Data, Linked Open Data, Veröffentlichung, Semantic Web, Strategien, Nationalbibliotheken, Projekte, Standards, Gremien, Normdaten, Culturegraph, Implementierungsszenarien, freie Lizenz, Österreich

LINKED (OPEN) DATA – BIBLIOGRAPHIC DATA ON THE SEMANTIC WEB. REPORT OF THE WORKING GROUP ON LINKED DATA TO THE PLENARY ASSEMBLY OF THE AUSTRIAN LIBRARY NETWORK

Abstract: *Linked Data stands for a certain approach to publishing data on the Web. The underlying idea is to harmonise heterogeneous data sources of different origin in order to improve their accessibility and interoperability, effectively making them queryable as a big distributed database.*

This report summarises relevant developments in Europe as well as the Linked Data Working Group's strategic and technical considerations regarding the publishing of the Austrian Library Network's (OBV's) bibliographic datasets. It concludes with the mutual agreement that the implementation of Linked Data principles within the OBV can only be taken into consideration accompanied by a discussion about the provision of the datasets under a free license.

Keywords: *linked data, linked open data, publishing, semantic web, strategies, national libraries, projects, standards, committees, standard data, Culturegraph, implementation scenarios, public domain, Austria*

Einleitung

Der Begriff „Linked Data“ steht für eine kleine Anzahl von Richtlinien zur Veröffentlichung von Daten im WWW (World Wide Web). Diese Richtlinien zielen insbesondere auf ein einheitliches Datenformat und „gemeinsames Vokabular“ der veröffentlichten Daten ab. Linked Data stellt somit eine Grundlage der Semantic Web¹-Vision dar, in der durch einheitliches Datenformat, eindeutige Bezeichner und taxonomisch geordnete, beschreibende Attribute die maschinelle Interpretation von ursprünglich heterogenen Daten unterschiedlichster Provenienz ermöglicht werden soll.

Die Entwicklungen der letzten Jahre im Bereich Linked Data haben auch vor dem Bibliothekssektor nicht haltgemacht. Parallel zu den Bestrebungen vieler europäischer Länder, Daten des öffentlichen Sektors im WWW zur Verfügung zu stellen – Stichwort „Open Government Data“ – haben in den letzten Jahren auch viele europäische Bibliotheken und Bibliotheksverbünde damit begonnen, ihre Daten – in diesem Bereich insbesondere Metadaten – in maschinell lesbarer Form zu veröffentlichen, sowie auch verfügbare externe Daten zur Kontextualisierung eigener Inhalte zu nutzen. Linked Data Prinzipien spielen in diesem Zusammenhang eine wichtige Rolle.

Dieser Bericht der von der Vollversammlung des Österreichischen Bibliothekenverbundes eingesetzten Arbeitsgruppe Linked Data fasst die Grund-

prinzipien dieses Standards zusammen, stellt einige herausragende Initiativen vor und präsentiert, dem Auftrag der Arbeitsgruppe entsprechend, mögliche Anwendungsfälle von Linked Data im Verbundumfeld.

1. Was ist Linked Data?

Die zentralen Richtlinien von Linked Data² wurden von Tim Berners-Lee, dem Vater des World Wide Web (WWW), formuliert und lauten sinngemäß:

1. Alles, wirklich alles, sollte durch eine URL (Uniform Resource Locator) identifiziert werden.
2. Um Dinge zu beschreiben, i.e. sie mit Metadaten zu versehen, sollte der Standard RDF (Resource Description Framework) verwendet werden.
3. Die so erstellen Daten sollten miteinander verbunden werden, damit Maschinen einfach Zusammenhänge herstellen können. Dies kann in erster Linie über gemeinsam genutzte, einheitliche Bezeichner und Vokabulare erreicht werden.

Eine URL ist eine Art der Identifikation, ein Uniform Resource Identifier (URI), von Ressourcen im Web. Ursprünglich wurde sie eingeführt, um Dokumente auf entfernten Rechnern eindeutig zu kennzeichnen und gleichzeitig zugänglich zu machen (z.B. <http://>-URL im WWW). Im Rahmen der Linked Data-Initiative wurde die Bedeutung von URLs erheblich erweitert, um auf diese Weise neben Dokumenten, den sogenannten „Information Resources“, auch physische Entitäten (natürliche Personen, Orte etc.), sogenannte „Non-Information Resources“, bezeichnen zu können. Die Verwendung von <http://>-URLs stellt sicher, dass diese Bezeichner auch als Web-Adresse fungieren können, über die Informationen über die bezeichnete Identität abgerufen werden können.

Informationen, die in Form von Linked Data bereitgestellt werden, sind in Form einfacher Sätze, sogenannter RDF-Triples, formuliert. Ein Triple besteht aus einem Subjekt, einem Prädikat und einem Objekt. Beispiel für Tripel sind: Die Sonne dreht sich um die Erde. Die Ampel zeigt Rot. Rot ist eine Farbe. In dieser Form ähneln Tripel den in der bibliothekarischen Welt bekannten Aussagen über Medien: Dieses Buch hat den Titel „Irrtümer der Menschheitsgeschichte“. Das Buch „Irrtümer der Menschheitsgeschichte“ ist im Jahr 2013 erschienen. Das Buch „Irrtümer der Menschheitsgeschich-

te“ ist ein gedrucktes Werk. Solche Aussagen werden heute i.d.R. im MAB bzw. MARC-Format codiert. Die Bedeutung der verwendeten Begriffe oder Codes ist meist in nur für Menschen lesbaren Dokumentationen festgehalten. RDF hingegen erlaubt eine Formalisierung. Jedes verwendete Subjekt (in diesem Fall ein Buch) kann einer Klasse, einer Kategorie von Dingen zugeordnet werden, ebenso das Prädikat und (meistens auch) das Objekt. Klassen und deren Beziehungen sind in sogenannten Ontologien festgehalten, die, einem Thesaurus nicht unähnlich, Beziehungen zwischen Dingen abbilden. Eine solche Kennzeichnung hilft einerseits dabei, bestimmte Daten aus einer großen Datenmenge gezielt zu extrahieren, ähnlich der Funktion einer Datenbank, aber darüber hinaus erlaubt sie es einem Computer beispielsweise zu folgern, dass das Buch nicht nur ein Druckwerk ist, sondern gleichzeitig zur übergeordneten Klasse der Schriftstücke gehört und dass Schriftstücke in der Regel von Menschen verfasst werden, nicht von Tieren oder Gegenständen. Solche Sachverhalte mögen zwar für Menschen selbstverständlich sein, müssen einer Maschine jedoch explizit mitgeteilt werden.

Die dritte – und zentrale – Forderung von Berners-Lee ist die nach der Verbindung von solcherart ausgezeichneten Daten. Ein entsprechendes Konzept ist aus Bibliothekskatalogen bekannt, die heute schon meist auf eine entfernte Normdatei, wie die Gemeinsame Normdatei (GND), oder zu anderen Quellen, wie Wikipedia, verlinken. Für Wikipedia existiert beispielsweise bereits eine RDF-Darstellung wesentlicher Fakten einzelner Artikel, die durch das Projekt DBpedia³ realisiert wird. Die GND wird von der Deutschen Nationalbibliothek ebenfalls in Form von RDF zum Download⁴ angeboten. Je mehr Institutionen ihre (Meta-)Daten in solcher Form bereitstellen, desto eher könnte der Idealfall eintreten, in dem jede Ressource im Web in strukturierter Form über viele Wege mit jeder anderen Ressource im Web verbunden ist und das WWW somit eine einzige, große Datenbank darstellt, die nicht wie derzeit üblich nur durch Menschen, sondern auch durch Maschinen erschließbar ist.

Abbildung 1 zeigt beispielsweise, wie vom Buch „Die Leiden des jungen Werthers“ von J.W.Goethe über die DBpedia-Ressource des Autors auf das Konzept „Sturm und Drang“ verknüpft werden kann, das seinerseits wiederum auf Friedrich Schiller verweist. Ebenfalls kann über Goethe zum Ort Weimar und darüber dann zum später dort verorteten Bauhaus verknüpft werden. Verknüpfungen wie diese könnten für neuartige Suchdienste sowie für wissenschaftliche Analysen gewinnbringend eingesetzt werden.

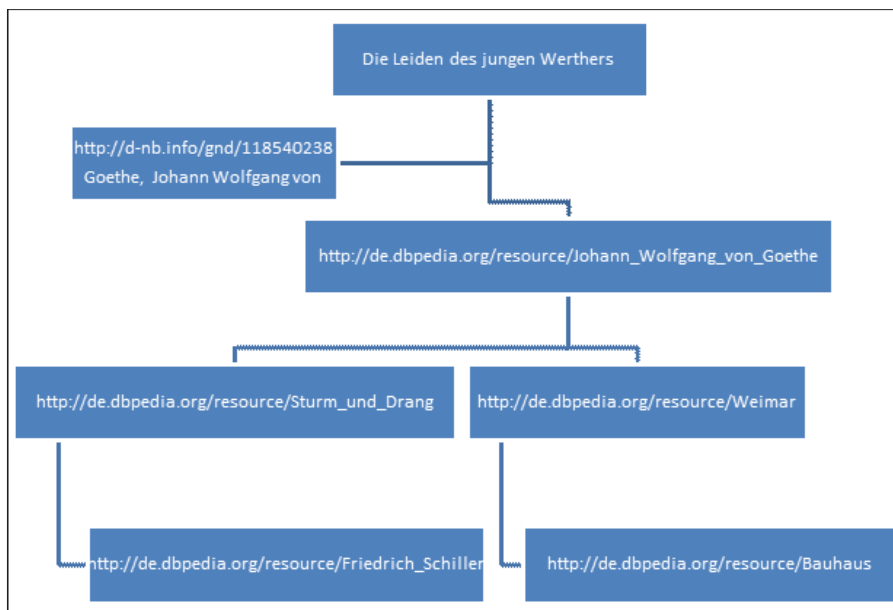


Abb. 1: Linked Data am Beispiel Johann Wolfgang von Goethe

Während mit dem Begriff Linked Data einerseits ein ganz bestimmtes Technologiebündel (URIs, Ontologien, RDF) gemeint ist, wird er andererseits auch verwendet, um generell die Verknüpfung heterogener Datensammlungen zu bezeichnen. In diesem Fall wird Linked Data mehr als Paradigma verstanden.

2. Linked Data (LD) und Linked Open Data (LOD)

Nachdem erste Datensätze als Linked Data veröffentlicht wurden, entstand die Diskussion, ob Linked Data überhaupt möglich ist, ohne dass diese Daten frei zur Verfügung stehen. Tim Berners-Lee bezog sehr schnell Stellung und überarbeitete seine Design Issues um den wichtigen Aspekt der Offenheit, die er unmittelbar mit der Frage von Linked Data verknüpft sieht. Rein technisch ist es jedoch möglich, Linked Data zu verwenden, ohne die Daten frei zur Verfügung zu stellen. Jedoch stellt Tim Berners-Lee selbst dieses Vorgehen in Frage, da diese Daten sich nur unter der Inkaufnahme einer Rechtsunsicherheit weiter verwenden lassen. Aus diesem Grund wurden inzwischen viele Daten unter einer freien Lizenz veröffentlicht⁵.

Lizenzen für Daten

Für Daten gilt, wie für jeden anderen „Content“ auch, dass bei der Veröffentlichung automatisch die im Urheberrecht festgeschriebenen Regeln gelten. Dies bedeutet für Daten, dass zwar der einzelne Datensatz selbst keinem Urheberrecht unterliegt, da dies Schöpfungshöhe hierfür nicht gegeben ist, jedoch gilt für eine Sammlung von Daten ein Datenbankschutz, da die Auswahl und Zusammenstellung der Daten eine schützenswerte Leistung darstellt. Aus diesem Grund ist die rechtliche Frage sehr kompliziert, wenn Daten ohne Angabe einer Lizenz als Linked Data zur Verfügung stehen. Insbesondere kann in diesem Fall davon ausgegangen werden, dass eine Verwendung des gesamten Abzugs der Daten (ein „Datenbank-Dump“) rechtlich nicht möglich ist, ohne geltendes Recht zu verletzen⁶. Aus diesem Grund sollten Datenabzüge mit einer Lizenz ausgestattet werden. Dafür bieten sich verschiedene Varianten an.

Doch welche Lizenzform empfiehlt sich für das Veröffentlichen von Daten? Im Jahr 2010 stellte Berners-Lee ein Sterne-System für Linked Open Data vor. Dabei bekommen alle Daten, die frei lizenziert sind, einen Stern. Zwei Sterne erhalten Daten, die zusätzlich noch in einer strukturierten Form (also Excel oder auch Endnote-Dateien) vorliegen. Wenn diese strukturierte Form noch ein offenes Format ist (wie beispielweise CSV oder auch MARC21) erhalten die Daten drei Sterne. Sobald die Daten mit Hilfe von W3C- Standards wie RDF oder SPARQL zur Verfügung gestellt werden, erhalten die Daten vier Sterne. Sobald diese Daten auch noch andere Datensätze als die eigenen verlinken, erhalten die Daten fünf Sterne. Voraussetzung ist jedoch immer, dass die Daten unter einer offenen Lizenz stehen. Nun stellt sich die Frage, bei welchen der bereits genannten Lizenzen handelt es sich um offene Lizenzen? Hierzu ist es wesentlich, dass der Term der Offenheit klar definiert ist. Offenheit bedeutet hier nicht, dass der Zugriff auf die Daten in einer Form kostenlos sein soll, sondern Offenheit ist hier im Sinne der Open Content Definition zu verstehen:

★ Available on the web (whatever format) but with an open licence, to be Open Data
★★ Available as machine-readable structured data (e.g. excel instead of image scan of a table)
★★★ as (2) plus non-proprietary format (e.g. CSV instead of excel)
★★★★ All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff
★★★★★ All the above, plus: Link your data to other people's data to provide context

Tab. 1: 5 Star System of Openness⁷

Theoretisch in Frage kommen alle Formen von Content Lizenzen wie beispielsweise die verschiedenen Creative Commons Lizenzen⁸. Jedoch eignen sich für das Bereitstellen von Daten in einer wirklich nachnutzbaren Form nur wenige Lizenzen. Wie auch im Bereich von Publikationen⁹ ist umstritten inwieweit die verschiedenen Creative Commons Lizenzen den Anforderungen der Offenheit genügen. Orientiert man sich an der Open Content Definition ist dies nur für die Creative Commons Attribution (Namensnennung – CC-BY) und die Creative Commons Attribution Share Alike (CC-BY-SA) der Fall. Für Daten ist aber selbst die Verwendung von Creative Commons aufgrund der Verpflichtung zur Namensnennung nicht optimal, da dies bei der Kombination von sehr vielen Datensätzen zu Problemen führen kann, dass bspw. die Anzeige von Ergebnissen kürzer ist als die Liste der verwendeten Quellen. Am universellsten lassen sich daher Daten einsetzen, für die keine Verpflichtungen bestehen, weil sie entweder Gemeingut (Public Domain) sind, oder weil der Urheber der Daten erklärt hat, dass er auf seine bestehenden Rechte verzichtet.

Es haben sich verschiedene Möglichkeiten etabliert, Daten als Gemeingut zur Verfügung zu stellen, dies ist nach deutschem Recht (sowie in vielen anderen europäischen Ländern) nicht mit Hilfe einer Lizenz möglich. Daher sind diese Varianten mehr als Verzichtserklärung anzusehen, von den verbleibenden Rechten jemals Gebrauch zu machen. Die gängigste Form einer solchen Erklärung stellt der Creative Commons 0 (CC0) Waiver dar. Der Waiver stellt im Europäischen Rechtsraum, in dem ein Verzicht auf Persönlichkeitsrechte nicht möglich ist, eine Erklärung dar, von den beim Urheber verbleibenden Rechten keinen Gebrauch zu machen. Da diese Form der Freigabe von Daten die weitestgehende Wiederverwendung erlaubt, wird diese von Creative Commons, der Open Knowledge Foundation und der DINI AG KIM Linked Library Data¹⁰ empfohlen. Verschiedene Bibliotheken, die bisher ihre Daten freigegeben haben, haben bereits den CC0 Waiver verwendet, beispielsweise das CERN, Bibliotheken aus dem hbz-Verbund und Bibliotheken des BVB und des KOBV (siehe auch Kapitel 4.2.).

3. Die Open Data Strategien der Nationalbibliotheken

Der Austausch von Metadaten hat in der Bibliothekswelt bereits seit langer Zeit Tradition. Weit verbreitete Standards wie die bereits weiter oben erwähnten Formate MARC und MAB gehen bis in die 1960er bzw. 1970er Jahre zurück. Diese Standards dienten ursprünglich dem Austausch von

Daten zwischen Bibliotheken bzw. innerhalb von Verbundsystemen, in der Regel innerhalb eines Landes oder Sprachraumes, um den Katalogisierungsprozess zu rationalisieren bzw. zu vereinheitlichen. Einige (National-) Bibliotheken stellten ihre Metadatenbestände, in der Regel Titeldaten, Inhaltsinformationen sowie Normdaten, auch Dritten zur Verfügung. Diese Dienste waren in der Regel kostenpflichtig und konnten im Rahmen von jährlichen Abonnements über elektronische Schnittstellen bezogen werden.

Die Zurverfügungstellung von Datenbeständen öffentlicher Institutionen wurde in den letzten Jahren immer mehr zum Fokus der Politik. Bereits im Jahr 2003 wurde von der EU die „Public-Sector-Information“ (PSI)-Richtlinie beschlossen und mit Stichtag 8.5.2008 von allen 27 Mitgliedsstaaten in nationales Recht überführt. Hierbei handelt es sich um die Schaffung eines einheitlichen rechtlichen Rahmens hinsichtlich der öffentlichen Verfügbarkeit von Inhalten des „öffentlichen Sektors“, der die Mitgliedsstaaten dazu ermutigen soll, so viel Information öffentlichen und damit auch von der Allgemeinheit finanzierten Ursprungs wie möglich für die Wiederverwendung verfügbar zu machen. Dies stützt sich im Wesentlichen auf zwei Säulen: Transparenz und fairen Wettbewerb. Die praktische Bedeutung dieser Richtlinie bezieht sich grundsätzlich darauf, wie öffentliche Institutionen auf Anfragen zur Freigabe von Datenbeständen reagieren müssen und umfasst unter anderem auch Höchstgrenzen für dafür einzuhebende Gebühren sowie das Gebot, die Daten allen Interessierten zu gleichen Konditionen anbieten zu müssen.¹¹

In ihrer ursprünglichen Form erstreckt sich diese Initiative ausdrücklich nicht auf die Sparten Bildung, Wissenschaft, Medien und den Kultursektor. Im Dezember 2011 wurde jedoch im Rahmen der EU Open Data Strategie¹² eine Erweiterung der Richtlinie angestrebt, die ausdrücklich auch den Kultursektor miteinschließt. Die Erweiterung befindet sich derzeit auf dem Weg zur Abstimmung im europäischen Parlament und es ist noch nicht ganz klar, wie weit sich eine allfällige Erstreckung der Richtlinie auf öffentliche Kulturinstitutionen auswirken wird. Es ist jedenfalls davon auszugehen, dass von Bibliotheken gepflegte Metadaten von dieser Richtlinie betroffen sein werden, nicht jedoch digitalisierte Inhalte¹³. Die Umsetzung dieser Erweiterung ist unter europäischen Kulturinstitutionen und Regierungen umstritten, beispielsweise hat die niederländische Regierung bereits ihre Ablehnung kundgetan. Auf der anderen Seite stehen zahlreiche Open-Data Initiativen, die diese Erweiterung befürworten, da sie einen großen Zuwachs an frei verfügbaren Daten verspricht.

Innerhalb der Europäischen Union stellt die „Digitale Agenda für Europa“¹⁴, die Strategie der EU, „digitale Technologien, einschließlich des Internet, dabei zu unterstützen, nachhaltiges Wirtschaftswachstum zu ermöglichen“, eine der Säulen der „Europa 2020“ Initiative dar. Die angestrebte Erweiterung der PSI-Richtlinie steht klar in diesem Zusammenhang. Eine weitere europäische Initiative, die im Rahmen der Digitalen Agenda als Priorität genannt ist, ist das gemeinsame europäische Kulturportal Europeana¹⁵.

Bei Europeana¹⁶ handelt es sich um die Initiative, dem von allen europäischen Kulturinstitutionen verwalteten Kulturgut eine einheitliche Internet-Plattform zu geben. Virtuelle BesucherInnen aus aller Welt sollen hier einheitlichen Zugang zu digitalen Repräsentationen der in europäischen Bibliotheken, Museen, Galerien und Archiven gesammelten Objekte erhalten. Dies soll einerseits dazu dienen, die Sichtbarkeit der europäischen Kulturinstitutionen zu erhöhen, indem „BesucherInnen“ der Europeana zum Betrachten einer digitalen Repräsentation auf die Web-Präsenz der jeweiligen Kulturinstitution weitergeleitet werden, andererseits aber auch durch die hier gesammelten Daten neue Erschließungs- sowie Verwertungsmöglichkeiten zu schaffen. Als Beispiel dafür sind etwa mobile Applikationen im Tourismusbereich zu nennen. Neuere Projekte im Rahmen der Europeana zielen jedoch auch darauf ab, direkten Zugang zu digitalen Objekten (z.B. Direktzugriff auf einzelne Seiten einer Zeitung anstatt Link auf Betrachter-Applikation) zu gewährleisten (auf freiwilliger Basis), wie es beispielsweise bei themenspezifischen Projekten wie Europeana Newspapers¹⁷ oder DM2E¹⁸ der Fall ist. In diesem Zusammenhang besteht auch das Ziel, eine verteilte Infrastruktur für die Forschung zu bieten, in der ForscherInnen institutionsübergreifender Zugang zu digitalen Inhalten gewährleistet wird.

Hinsichtlich der Verfügbarkeit der in Europeana aggregierten Metadaten gilt seit 1.7.2012 das neue Europeana Data Exchange Agreement (DEA)¹⁹, in dem sich alle Institutionen, die der Europeana Daten zur Verfügung stellen, verpflichten, ihre Metadaten unter der Creative Commons CC0 1.0 Universal Public Domain Dedication (s.o.) zur Verfügung zu stellen, soweit dies nach lokalem Recht möglich ist – bei Unvereinbarkeit existiert ein sogenannter „Public License Fallback“²⁰. Diese Entwicklung folgt damit der weiter oben erwähnten EU Open Data Strategie und der Digitalen Agenda für Europa.

In diesem Zusammenhang ist auch die Erklärung der CENL²¹ (Foundation Conference of European National Libraries) vom 28.11.2011 zu sehen, in

der ihre 49 Mitglieder, darunter auch die Österreichische Nationalbibliothek, sich auf Basis einer deutlichen Abstimmungsmehrheit darauf einigen, die von ihnen im Rahmen der Europeana zur Verfügung gestellten Daten unter CC0 anzubieten. Die Situation bezüglich der CC0 Lizenz stellt sich in Österreich grundsätzlich so dar, dass Urheberpersönlichkeitsrechte nicht verzichtbar sind und daher in solchen Fällen die CC0 Lizenz nicht anwendbar ist. Wie weiter oben bereits erwähnt, stellen jedoch die von Bibliotheken bzw. dem Verbund zur Verfügung gestellten Metadaten nach Regelwerk erstellte Daten dar, die als solche keinen urheberrechtlichen Schutz genießen, sondern bloß – in ihrer Gesamtheit – einen Investitionsschutz, für den sich die Frage der unverzichtbaren Urheberpersönlichkeitsrechte nicht stellt.

Das gemeinsame Bekenntnis der europäischen Nationalbibliotheken zu Open Data stellt den Gipfel einer längeren Entwicklung dar. Betrachtet man diese im europäischen Raum, so wird deutlich, dass im Laufe der letzten Jahre eine Reihe von europäischen Nationalbibliotheken damit begonnen haben, Teile ihrer (Katalog-)Daten, in der Regel die jeweiligen Nationalbibliographien, in Form von LOD anzubieten. Abbildung 2 zeigt diese Entwicklung anhand von sieben Beispielen, auf die in Abschnitt 4.1. näher eingegangen wird.

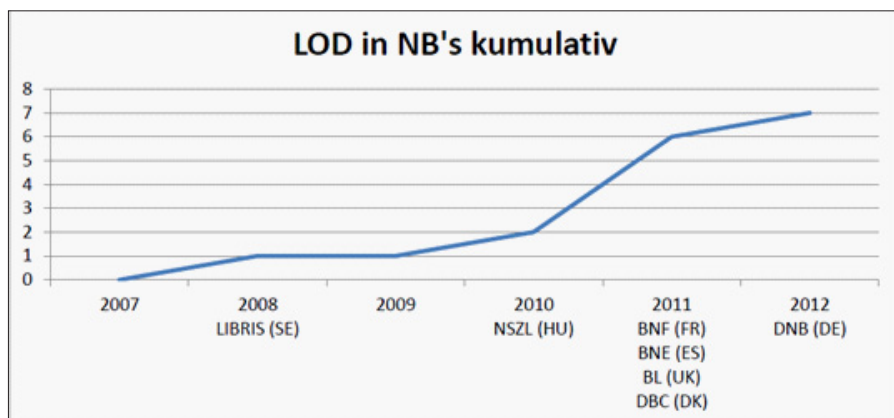


Abb. 2: Europäische Nationalbibliotheken mit LOD Angebot

Motive

Die institutionellen Motive für die Bereitstellung von Daten via LOD sind vielfältig. Abgesehen von technischen Überlegungen, wie Bibliothekskata-

loge am „natürlichsten“ in das WWW integriert werden können²², stehen auch strategische Erwägungen im Vordergrund.

In einem Paper zur IFLA 2012²³ beschreibt die Generaldirektorin der Deutschen Nationalbibliothek (DNB), Elisabeth Niggemann, die gegenwärtige Situation sinngemäß als „Konkurrenz“ zwischen den traditionellen Produzenten von Metadaten, den Bibliotheken und Verlagen, und den wachsenden Metadatenbeständen, die im Rahmen des E-Commerce (Amazon) und der rasant wachsenden Verfügbarkeit an elektronischen Publikationen frei im Internet zugänglich sind. Sie unterstreicht die Tatsache, dass die wertvollen Metadaten der Bibliotheken im Internet derzeit nur schwer zu finden sind und für Bibliotheken daher die Notwendigkeit besteht, ihre Sichtbarkeit im Internet zu erhöhen, wenn sie ihre führende Rolle in der Wissensorganisation beibehalten wollen. Sie führt darüber hinaus an, dass urheberrechtliche Bedenken bei Metadaten in der Regel keine Rolle spielen und das Zögern von Bibliotheken, Archiven und Museen, Metadaten frei zur Verfügung zu stellen, daher eher einer Angst vor dem Verlust der Kontrolle über diese Daten entspringt. Sie unterstreicht auch die Rolle, die das oben erwähnte Europeana Data Exchange Agreement (DEA) und die Vorbildfunktionen von Institutionen wie der British Library oder der Spanischen Nationalbibliothek bei der Entscheidung der DNB gespielt haben, sowohl die Deutsche Nationalbibliographie als auch die Gemeinsame Normdatei in Form von LOD zur Verfügung zu stellen.

Auf der selben Konferenz erklärt Gildas Illien, Direktor der Abteilung für bibliographische und digitale Information an der Französischen Nationalbibliothek (BNF), dass – so wie für die DNB – das Europeana DEA, aber auch die Eröffnung des nationalen französischen Datenportales data.gouv.fr wesentlich zur Entscheidung der BNF, Katalogdaten als LOD anzubieten, beigetragen haben²⁴. Er betont den Umstand, dass „Linked Data“ allein nicht ausreicht, sondern nur „Linked Open Data“ Sinn ergibt, da nur auf diese Weise zusätzlich zur technischen/inhaltlichen auch die rechtliche Interoperabilität zwischen den Datensätzen verschiedener Institutionen gewährleistet werden kann.

Wie die beiden obigen Beispiele zeigen, haben sowohl das Europeana Data Exchange Agreement als auch allgemeine Entwicklungen im Bereich Open Government Data Einfluss auf die Open Data Strategien der europäischen Nationalbibliotheken und es ist zu erwarten, dass sich dieser Trend weiterhin fortsetzen wird. Mittlerweile erscheinen auch erste Erfahrungsberichte

über die Akzeptanz von existierenden LOD Angeboten. In einem Paper zur „Open Data on The Web“ Konferenz vom April 2013 beschreibt Lars Svensson von der deutschen Nationalbibliothek, dass der Schritt der DNB zu CC0 sehr positives Echo erzeugt und hohe öffentliche Aufmerksamkeit gebracht hat²⁵. Als Nachteil merkt er an, dass sich durch die Öffnung der Datenbestände nur noch schwer verfolgen lässt, wie die freigegebenen Daten genutzt werden und nur gelegentliche Rückfragen von BenutzerInnen vereinzelt darüber Aufschluss geben können, also nur „die Spitze des Eisberges“ der Datennutzung bekannt ist. Obwohl die Freigabe der Daten auch einen Rückgang der durch die zuvor kostenpflichtigen Datendienste erzielten Einnahmen bedeutet, schließt er jedoch damit, dass der Markt für Bibliotheksdaten in Zukunft mit hoher Wahrscheinlichkeit ohnehin schwinden wird und sich durch die freie Verfügbarkeit der Daten dafür neue Anwendungsszenarien ergeben, die einen Mehrwert bringen können.

4. Projekte

4.1. Konkrete Projekte in Europäischen Nationalbibliotheken

Wie bereits zuvor erwähnt, zeigt Abbildung 2 die Zunahme des LOD Angebotes der europäischen Nationalbibliotheken der letzten Jahre. Die angebotenen Inhalte und Dienste unterscheiden sich jedoch teilweise deutlich voneinander, daher wird hier kurz auf die verschiedenen Realisierungen eingegangen.

Linked Data im Bibliotheksbereich bezieht sich derzeit in erster Linie auf Katalogdaten, also Titel- und Normdaten. Entsprechende Angebote können unterschieden werden nach

a. den angebotenen Inhalten

Welche Lizenz gilt für die Nutzung der Daten?

Wird der komplette Bibliothekskatalog als LOD veröffentlicht, oder nur Teile davon?

Bezieht die Veröffentlichung nur Titel-, oder auch Normdaten mit ein?

Welches Datenmodell/Vokabular wird eingesetzt, um die Katalogdaten abzubilden?

b. den angebotenen Diensten

Wie kann man die Inhalte als LOD abrufen?

Als einzelne Titeldaten, als Komplett-Download oder über eine Datenbankabfragesprache?

Wie oft werden die angebotenen Inhalte aktualisiert?

Es wurden die Angebote der folgenden sieben Nationalbibliotheken untersucht. LIBRIS und DBC stellen hier einen Sonderfall dar, da es sich bei LIBRIS um einen Zugang zum schwedischen Verbundkatalog handelt, der von Schwedischen Nationalbibliothek verwaltet wird, und im Falle von DBC es sich um eine eigene, vom Dänischen Staat beauftragte Institution handelt, die gemeinsam mit der Dänischen Nationalbibliothek die dänische Nationalbibliographie herausgibt:

Name	Land	Jahr	URL
LIBRIS	Schweden	2008	http://librisbloggen.kb.se/2008/12/03/libris-available-as-linked-data/
NSZL	Ungarn	2010	http://nektar.oszk.hu/wiki/Semantic_web
BNF	Frankreich	2011	http://data.bnf.fr/about-en
BNE	Spanien	2011	http://datos.bne.es
BL	Großbritannien	2011	http://www.bl.uk/bibliographic/datafree.html
DBC	Dänemark	2011	http://oss.dbc.dk/plone/linked-data
DNB	Deutschland	2012	http://www.dnb.de/DE/Service/DigitaleDienste/LinkedData/linkedata_node.html

Inhalte

Lizenz

Die in Kapitel 2 und 3 erwähnten Gründe haben die meisten Institutionen dazu bewogen, ihre Daten unter einer offenen Lizenz anzubieten. Fünf von sieben Institutionen bieten ihre Metadaten unter CC0 an (LIBRIS erst ab 2011), während die Lizenz der BNF²⁶ eine dem französischen Recht angepasste Version der CC-BY 2.0 darstellt und die Ungarische Nationalbibliothek keinen Hinweis auf eine Lizenz gibt.

Titel- und Normdaten

Das Angebot umfasst bei drei Bibliotheken (LIBRIS, BL, DNB) die jeweilige Nationalbibliographie, bei der NSZL den kompletten OPAC sowie die gesamte Digitale Bibliothek. Die BNE bietet eine nicht genauer spezifizierte Teilmenge ihres Kataloges in Form von 2.400.000 Titeldaten an, während

die BNF einen ausgewählten Satz von derzeit 3.600.000 Titeldaten bereitstellt. DBC unterhält derzeit nur einen Testbetrieb mit der dänischen Nationalbibliographie ab 2010. Alle Nationalbibliotheken bieten auch komplementär zu den Titeldaten die entsprechenden Normdaten an.

Datenmodellierung

Deutliche Unterschiede existieren in der Modellierung der Daten. Allgemein kann jedoch gesagt werden, dass die meisten Angebote versuchen, bereits etablierte Vokabulare und Modelle einzusetzen. Während die spanische BNE und die französische BNF ihre Daten auf FRBR (Functional Requirements for Bibliographic Records) aufbauen, setzen die anderen Nationalbibliotheken eher auf eigene Modelle. Die British Library hat kürzlich das „BL Data Model“²⁷ veröffentlicht, dieses wird auch von der dänischen DBC in abgewandelter Form eingesetzt. Die Deutsche Nationalbibliothek setzt für ihre Normdaten die eigene GND-Ontologie ein. Bezüglich der eingesetzten Vokabulare/Ontologien dominieren bei den meisten betrachteten Bibliotheken Bibliographic Ontology (BIBO²⁸) und Dublin Core (DC²⁹) für Titeldaten sowie Friend-of-a-Friend (FOAF³⁰) für Personen und das Simple-Knowledge-Organisation-Schema (SKOS³¹) für Schlagworte.

Verknüpfung mit externen Datenquellen

Einer der Hauptvorteile des Bereitstellens von LOD ist die Verknüpfung der eigenen Datenbestände mit jenen von anderen Institutionen oder mit fachfremden Quellen, die der Kontextualisierung der eigenen Daten dienen können. Alle betrachteten Bibliotheken verknüpfen Personendaten mit dem Virtual-Internet-Authority-File (VIAF³²) oder können über ihre eigenen Bezeichner damit verknüpft werden. Auch verweisen alle Bibliotheken bis auf BL und DBC bei Personen und Schlagworten zu DBpedia/Wikipedia-Inhalten. Darüber hinaus wird auch oft auf Daten der Library of Congress (LOC, LCSH³³) verwiesen. Auch die GND wird von mehreren Bibliotheken als externe Verknüpfung eingesetzt.

Dienste

Es existieren vier grundlegende Möglichkeiten, LOD anzubieten. Darunter fällt:

a. Abruf von Datensätzen pro Titel/Person etc.

Diese Art des Zugriffes kann am ehesten mit dem Navigieren einer Person in einem Online-Katalog verglichen werden. Anhand eines eindeutigen Bezeichners wird zuerst ein Titel- oder Norm-

datensatz abgerufen und entlang der in diesem Datensatz vorkommenden Verknüpfungen dann entsprechend zu anderen Datensätzen gewechselt. In diesem Fall kommt auch oft sogenannte „Content Negotiation“ zum Einsatz, wobei ein und dieselbe URL je nach Anfrage entweder HTML für einen menschlichen Besucher oder RDF für eine maschinelle Abfrage zurückliefert. Der Vorteil dieses Ansatzes ist die Verfügbarkeit der Daten in Echtzeit, Nachteil ist, dass beim Folgen der Verknüpfungen im aktuellen Datensatz immer wieder ein neuer Datenabruf für nachzuladende Datensätze abgesetzt werden muss. Für den Download kompletter Datenbestände ist diese Variante daher nicht geeignet.

Diese Variante wird von allen betrachteten Nationalbibliotheken angeboten, wobei sie zum Zeitpunkt der Erstellung dieses Berichts bei NSZL und bei DBC gerade nicht funktioniert hat.

b. Abruf bzw. Download des kompletten Datensatzes auf den eigenen Rechner („Dump“)

Alle Datensätze werden in Form von einer oder mehreren Dateien auf der Website der Institution zum Download bereitgestellt. Der Vorteil ist, dass für z.B. Forschungszwecke alle Daten auf einmal für Analysen zur Verfügung stehen und dass seitens der bereitstellenden Institution nur wenige Anforderungen an die Infrastruktur bestehen. Der Nachteil ist, dass die Erzeugung der Download-Dateien nur in (un-)regelmäßigen Abständen geschieht und die Daten daher nicht immer auf dem neuesten Stand sind. Darüber hinaus kann es aufgrund der großen Menge an Datensätzen zu längeren Downloadzeiten kommen.

Diese Variante wird von allen Bibliotheken außer LIBRIS und NSZL angeboten

c. Zugriff via Datenbankabfrage mit der RDF-Abfragesprache SPARQL

Diese Zugriffsmöglichkeit bietet die größte Flexibilität für AnwenderInnen der Daten. Über die standardisierte Abfragesprache SPARQL kann bei entsprechenden Kenntnissen des zugrundeliegenden Datenmodells genau jene Teilmenge der verfügbaren Daten abgerufen werden, die gerade benötigt wird. Darüber hinaus bietet SPARQL auch eine Reihe von den von der relationalen Datenbankabfragesprache SQL bekannten Aggregatfunktionen, um beispielsweise

se alle Datensätze zu zählen, die ein bestimmtes Attribut aufweisen, ohne alle Datensätze herunterladen zu müssen. Der Nachteil ist der hohe Ressourcenaufwand auf Seite des Diensteanbieters, da komplizierte Abfragen sehr hohen Rechenaufwand verursachen können und AnwenderInnen über die entsprechenden Kenntnisse verfügen müssen, Abfragen entsprechend zu optimieren. Aus diesem Grund sind die SPARQL Schnittstellen vieler Institutionen dahingehend beschränkt, dass nur Abfragen, die eine gewisse Verarbeitungszeit nicht überschreiten, zulässig sind oder auch die Menge an zurückgegebenen Datensätzen limitiert und damit das Ergebnis einer Abfrage unter Umständen entsprechend unvollständig ist.

Diese Variante wird von allen Bibliotheken außer BNF und DNB angeboten.

d. Zugriff via Linked Data API

Diese Variante vereint die Vorteile von Punkt a) und Punkt c), in dem bestimmte Datenbankabfragen hinter einer eigenen Datenschnittstelle gekapselt werden. AnwenderInnen können ohne weitgehende SPARQL Kenntnisse selektive Datenabfragen durchführen. Dies hängt jedoch in erster Linie davon ab, welcher Aufwand bei der Entwicklung dieser Schnittstelle einfließt.

Diese Variante wird derzeit nur von BL angeboten.

4.2. Projekte in den deutschen Verbünden

Die vorgestellten Projekte maßgeblicher Nationalbibliotheken stehen exemplarisch für andere ähnlich gelagerte Unternehmungen mit dem Ziel, bibliographische Titel- und Normdaten gemäß den LD-Prinzipien zu veröffentlichen. Insbesondere im deutschsprachigen Raum gibt es kaum mehr einen großen Verbund, der dies nicht entweder schon in der ein oder anderen Form getan, oder entsprechende Pläne dazu hat: Eine Vorreiterrolle hat das Hochschulbibliothekszentrum Nordrhein-Westfalen (hbz) eingenommen und bereits im Jahr 2009 die Titeldaten zahlreicher Verbundteilnehmer zum Download bereitgestellt. Seitdem haben der B3Kat und HeBIS gleichgezogen.

hbz³⁴

Bereits im März 2010 setzten Kölner Bibliotheken und das Landesbibliothekszentrum Rheinland-Pfalz (LBZ) in Kooperation mit dem hbz den

Open-Access-Gedanken auch für bibliographische Daten um. Die Kataloge der Bibliotheken wurden der Öffentlichkeit zur freien (Nach-)Nutzung zur Verfügung gestellt, um die Vernetzung mit Daten aus anderen Bereichen im Semantic Web zu ermöglichen³⁵. Mittlerweile haben alle hbz-Verbundbibliotheken (bis auf zwei) einer Veröffentlichung der Katalogdaten ihrer Bestände unter CC0 zugestimmt³⁶. Mit <http://lobid.org> (lobid als Akronym für Linking Open Bibliographic Data) entstanden zunächst zwei experimentelle Dienste: lobid-resources (Metadaten als LOD bereitstellen) und lobid-organisations (internationales LD-basiertes Verzeichnis von Bibliotheken und verwandten Institutionen).

Der Gesamtbestand³⁷ kann heruntergeladen werden; darüber hinaus gibt es einen SPARQL-Endpoint.

B3Kat³⁸

Der Verbundkatalog B3Kat der Bayerischen Staatsbibliothek, des Bibliotheksverbunds Bayern und des Kooperativen Bibliotheksverbunds Berlin-Brandenburg haben ihren Katalog im Dezember 2011 als LOD veröffentlicht.

Der Gesamtbestand³⁹ kann heruntergeladen werden; darüber hinaus gibt es einen SPARQL-Endpoint.

HeBIS⁴⁰

Der HeBIS-Verbund (Hessisches Bibliotheks- und Informationssystem) hat die Katalogdaten seines Verbundes im Juli 2012 als LOD veröffentlicht.

Der Gesamtbestand⁴¹ kann heruntergeladen werden.

Im Gemeinsamen Bibliotheksverbund (GBV) wurde Ende 2012 die LOD-Publikation der bibliographischen Daten von der GBV-Verbundleitung genehmigt. Das Bibliotheksservice-Zentrum Baden-Württemberg hat Anfang des Jahres 2013 die Veröffentlichung der bibliographischen Daten des Südwestdeutschen Bibliotheksverbundes (SWB) beschlossen.

5. Bibliothekarische Standards und Gremien

Wie eingangs beschrieben, werden im Semantic Web nicht nur Dinge eindeutig über URIs identifiziert, sondern auch verwendete Vokabularien bzw. Relationen. Die Veröffentlichung von Titeldaten wird deswegen durch die Veröffentlichung der verwendeten Metadatenstandards ergänzt. War die Bedeutung der verwendeten beschreibenden Attribute wie „Titel“, „Au-

tor“, „Übergeordneter Satz“ etc. bis jetzt lediglich in Form schriftlicher Dokumentation vorhanden, müssen diese nun in ihrer Bedeutung in maschinell interpretierbarer Form festgelegt werden. Entsprechende Projekte sind in Umfang und Ausmaß noch nicht so weit gediehen wie die Veröffentlichung von Titeldaten – in vielen Fällen liegen bereits Vokabularien vor, welche, den LD Best Practices folgend, nachgenutzt werden. Es sind dies, wie in Kapitel 4.1. bereits erwähnt, Dublin Core, BIBO bzw. die Begriffe der RDA. Dennoch: Um besondere Relationen abbilden zu können musste z.B. die Deutsche Nationalbibliothek eine eigene GND-Ontologie entwerfen⁴². Ähnlich gelagerte Unternehmungen zur Veröffentlichung spezialisierter Ontologien für die Darstellung von Bibliotheksbeständen sind derzeit noch unvollendet (GBV, BIBFRAME). Die Bereitstellung von Titel- und Normdaten sowie der verwendeten Vokabularien wird begleitet, überwacht und angetrieben von einer Vielzahl von Arbeitsgruppen und Expertengremien auf nationaler und internationaler Ebene:

5.1. DCMI Bibliographic Metadata Task Group

Im Jahr 2007 fand das sogenannte „London Meeting“ (Data Model Meeting)⁴³ statt. Ziel dieses Meetings von Metadaten-ExpertInnen war es, die möglichen Verbindungen des neuen Content Standard Resource Description and Access (RDA) mit Modellen anderer Metadaten Communities, die zum Teil parallel entwickelt wurden, zu beleuchten. Die DCMI/RDA Task Group wurde gegründet, die sich aus Mitgliedern des Joint Steering Committees for Development of RDA (JSC), der Dublin Core Metadata Initiative (DCMI) und anderen Semantic Web-Gruppen zusammensetzte. Gemeinsam wurde an einer maschinenlesbaren Ontologie der RDA-Elemente gearbeitet - RDA Element Sets und RDA Vocabularies sind von dieser Gruppe unter <http://rdvocab.info/> erarbeitet worden und können für Linked Data-Anwendungen nachgenutzt werden. Fünf Jahre nach dem London Meeting fand erneut ein Treffen von Metadaten-ExpertInnen in London statt - das Seminar „Five Years On“⁴⁴. Die DCMI/RDA Task Group wurde reorganisiert und umbenannt in die „Bibliographic Metadata Task Group“⁴⁵.

5.2. Bibliographic Framework Transition Initiative – BIBFRAME

Bereits um die Jahrtausendwende wurde Kritik am Datenformat MARC21 laut, da es nicht mehr zeitgemäß schien, um die Metadaten-Bedürfnisse von Bibliotheken abzudecken. Und auch der US RDA Test im Jahr 2011 be-

stätigte, dass MARC21 nicht optimal mit RDA kompatibel scheint. Somit wurde von der Library of Congress die „Bibliographic Framework Transition Initiative“ (kurz: BIBFRAME) ins Leben gerufen - eine Initiative, die die MARC-Ablöse planen und begleiten soll. Die Grundsätze für das Nachfolgeformat wurden zu Beginn der Initiative im Mai 2011 klar dargelegt: MARC-Data als Linked Data und die Ausrichtung an Internet- und Linked Data-Umgebungen haben höchste Priorität. Als „Early Implementers“ sind neben der Library of Congress folgende Bibliotheken und Institutionen beteiligt: British Library, Deutsche Nationalbibliothek, George Washington University, National Library of Medicine, OCLC und Princeton University. Darüber hinaus wurde die Firma Zepheira engagiert, um ein Linked Data-Modell zu erarbeiten. Im November 2012 wurde ein erstes Papier mit dem Titel „Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services“⁴⁶ veröffentlicht. Enthalten ist ein erstes BIBFRAME-Modell, welches von der Bibliothekscommunity kommentiert werden kann und von den Early Implementers getestet wird.

5.3. DINI AG KIM

Das Kompetenzzentrum Interoperable Metadaten⁴⁷ ist eine Arbeitsgruppe der Deutschen Initiative für Netzwerkinformation (DINI) und vernetzt ExpertInnen im Bereich Metadaten und Semantic Web. Neben dem Wissenstransfer innerhalb der Gruppe ist KIM auch eine Plattform, um neue Projekte zu initiieren und fördert/unterstützt die Erarbeitung von Metadatenstandards sowie die interoperable Gestaltung von Formaten. Innerhalb von KIM haben sich mehrere Gruppierungen etabliert, die sich mit unterschiedlichen Themenschwerpunkten beschäftigen. Eine dieser Gruppierungen, nämlich die Titeldaten-Gruppe, hat sich zum Ziel gesetzt, eine Harmonisierung der RDF-Repräsentationen von Titeldaten zu erreichen und darüber hinaus eine Empfehlung hierzu zu erarbeiten. Linked Data-VertreterInnen der deutschsprachigen Verbünde und einiger Bibliotheken beobachten in dieser Gruppe aktuelle internationale Entwicklungen hinsichtlich Bereitstellung bibliographischer Daten im Semantic Web und möchten Erfahrungen und Anforderungen der deutschsprachigen Bibliothekswelt einbringen. Ein erster Entwurf der Empfehlungen zur RDF-Repräsentation bibliographischer Daten (vorerst von Textressourcen) ist als Version 0.1 im Wiki der Titeldaten Gruppe zugänglich⁴⁸. Moderation und Koordination dieser Gruppe liegt bei der Deutschen Nationalbibliothek und seit Dezember 2012 ist auch die OBVSG Mitglied der DINI AG KIM Titeldaten Gruppe.

6. Normdaten und Culturegraph

Ein dritter Bereich ergänzt jenen der Titeldaten und Vokabularien: die Veröffentlichung von Normdaten, welche zeitlich betrachtet sogar am Beginn der Library Linked Data-Ära standen. Als Vorreiter trat die Library of Congress auf, welche bereits im Jahr 2008 Vorschläge für die Transformation der Library of Congress Subject Headings erarbeitete⁴⁹. Viele Nationalbibliotheken haben seitdem Anstrengungen zur Veröffentlichung ihrer Normdaten in maschinenlesbarem Format unternommen, darunter die British Library, die Bibliothèque National de France und die Biblioteca Nacional de Espana. Auch die für Österreich besonders relevante Gemeinsame Normdatei ist inzwischen in ihrer Gesamtheit und in Echtzeit als RDF abrufbar. Gemäß den Linked Data Prinzipien kann über eine stabile URL, beispielsweise die URL <http://d-nb.info/118575147/about/rdf> für den Normdatensatz des deutschen Soziologen Niklas Luhmann, auf eine RDF-Version zugegriffen werden. Über die GND-Nummer lassen sich Verknüpfungen zur englischen und deutschen Wikipedia/DBpedia, sowie auch zur Internationalen Normdatei VIAF⁵⁰ herstellen. Insbesondere DBpedia, als die vielleicht wichtigste Linked Data-Ressource im gesamten Web, stellt den idealen Einstiegspunkt dar, um weitere, eventuell relevante Ressourcen zu erreichen. Die Daten wurden unter CC0 veröffentlicht. Auch die Kategorien bekannter, aber auch sehr spezialisierter, Klassifikationssysteme sind inzwischen gemäß den Linked Data-Prinzipien zugänglich: Die Dewey Dezimalklassifikation beispielsweise ist bis zu drei Ebenen tief in maschinenlesbarem Format mehrsprachig abrufbar⁵¹. Die im deutschen Sprachraum sehr verbreiteten Klassifikationen „Regensburger Verbundklassifikation“ und „Basisklassifikation“ hingegen sind es nach wie vor nicht.

Ein Projekt, das keinem der drei genannten Bereiche eindeutig zuordenbar ist, ist culturegraph.org. Culturegraph.org wurde gemeinsam von der Deutschen Nationalbibliothek und dem hbz ins Leben gerufen und sollte ursprünglich der unkontrollierten Vervielfachung von URLs für Bibliotheksmedien entgegenwirken. Denn wenngleich es jedem frei steht, neue URLs für im eigenen Katalog verzeichnete Bücher zu prägen, wie z.B. permalink.obvsg.at/AC12345678, könnte diese Praxis dazu führen, dass die so veröffentlichten Datenbestände unverbunden und isoliert nebeneinander stehen. Culturegraph.org hat dementsprechend den Anspruch, verbundübergreifende URLs zu prägen, welche jeweils zu den lokalen Identifiern in den Verbünden verweisen und umgekehrt. Diese Vorgangsweise ist vergleichbar mit der Zusammenführung von Personennormdaten im Rahmen der

VIAF-Initiative. Um solche Verweisungs-bündel zu erstellen, hat der OBV im Rahmen einer AG KVA-Initiative bereits Daten an die DNB geliefert. Die entstandenen Konkordanzen können natürlich nicht nur für Linked Data-Anwendungen, sondern auch für klassische Kataloganreicherungsprojekte nachgenutzt werden. Erst kürzlich wurde von Seiten der Projektinitiatoren ein Linked Data-Service in Aussicht gestellt, welches in erster Linie für die Deutsche Digitale Bibliothek⁵² entwickelt wird. Sollte sich culturegraph.org so dynamisch entwickeln, wie es derzeit den Anschein hat, dann könnte es zu einem zentralen Dreh- und Angelpunkt zwischen der Bibliothekswelt und anderen Datenbeständen im Semantic Web werden. Eine vollständige Teilnahme des OBV an diesem Projekt würde jedoch voraussetzen, daß die OBV-Metadaten unter einer offenen Lizenz bereitgestellt werden, da alle in Culturegraph.org gesammelten Daten schließlich in kombinierter Form verfügbar gemacht werden sollen und dies nur unter einer einheitlichen, offenen Lizenz funktionieren würde.

7. Firmenpolitik

Die derzeit größten Anbieter von Bibliothekssystemen – OCLC und ExLibris – sind in unterschiedlicher Ausprägung auf dem Linked Data-Sektor aktiv. Im Folgenden der Versuch einer kurzen Zusammenfassung.

OCLC – am österreichischen „Markt“ nicht wirklich vertreten – kann derzeit wohl als der weltgrößte Anbieter von bibliothekarischen Daten – Titel- und Bestandsdaten gelten. Die zum Teil seltsame und rechtlich recht restriktive Politik in Bezug auf die Nutzung der im WorldCat versammelten Daten erhielt im vergangenen Jahr eine neue Dynamik mit der Entscheidung von OCLC, mehr in den Bereich Linked (Open) Data zu investieren. Mit der Ernennung von Richard Wallis zum Technology Evangelist bei OCLC wurde ein Spezialist gefunden, der zuvor bei Talis, einer in Großbritannien führenden Firma im Bereich von Linked Data und Semantic Web-Technologien, gearbeitet hat. Seine Aufgabe bei OCLC wird es sein, neue Internettechnologien (also Sematische Webtechnologien) zu implementieren. Seit dem Jahr 2012 hat sich in diesem Bereich bei OCLC einiges getan. Dabei hat man sich bei OCLC für eine sehr freie Lizenz (ODC-BY) entschieden. Zudem wurde die gesamte 23. Ausgabe der Dewey Decimal Classification 2012 als Linked Data veröffentlicht. Darüber hinaus wurden zu jedem bibliographischen Datensatz in World-Cat als erstem Schritt in Richtung Linked Data schema.org folgende de-

skriptive Auszeichnungen hinzugefügt. OCLC bezeichnet dies selbst als nur den ersten Schritt.

Der für Österreich wichtige Produzent bibliothekarischer Anwendungen, ExLibris, signalisiert ebenso Interesse am Einsatz von semantischen Technologien und engagiert sich in einer Arbeitsgruppe zur Verbesserung der maschinenlesbaren Angaben in Bibliothekskatalogen⁵³. Außerdem wird ein intensiver Kontakt mit der Linked Data-Interessensgemeinschaft der IGeLU⁵⁴ gesucht, welche ihrerseits Vorschläge für den Einsatz von Linked Data Technologien in Ex Libris Produkten erarbeitet. Für Aleph wird eine technologische Umrüstung auf Linked Data-Standards verständlicherweise nicht mehr angedacht und während die Perspektive für ALMA noch unklar ist, wird Primo, als dem sog. Discovery-Layer, eine Schlüsselrolle zukommen. Erste konkrete Veränderungen der Software in diese Richtung sind mit dem jüngsten Systemupdate erfolgt.

8. Implementierungsszenarien & Anwendungsfälle in Österreich

Angesichts der Fülle von Projekten und Vorhaben, wie sie auszugsweise oben vorgestellt wurden, kann die bisher entfaltete Library Linked Data-Aktivität in Österreich als bescheiden bezeichnet werden. An internationalen Kooperationen teilnehmende Einrichtungen, wie die Österreichische Nationalbibliothek oder die Universität Wien, sehen sich zwar mit solchen Anforderungen konfrontiert, einen österreichweiten, oder auch nur verbundweiten Konsens über die Möglichkeiten des Einsatzes von Linked Data-Technologien bzw. deren Sinnhaftigkeit gibt es aber bis dato nicht.

Einer Einteilung der Linked Data-Arbeitsgruppe der IGeLU folgend, unterscheiden wir grundsätzlich zwischen Projekten, die sich bereits vorhandene Linked Data-Quellen zu Nutze machen, um frei verfügbare Daten zu konsumieren, und solche, die mit der Veröffentlichung eigener Daten zu tun haben. Während der OBV bereits frei verfügbare Daten wie etwa DBpedia benutzt, spielte die Auseinandersetzung mit der Veröffentlichung der eigenen Daten bisher eine untergeordnete Rolle. Abseits der Bibliothekswelt kann in den letzten Monaten aber auch in Österreich eine zunehmende Diskussion über die Zurverfügungstellung öffentlicher Daten beobachtet werden⁵⁵, die längerfristig, auch im Zusammenhang mit den in Abschnitt 3 erwähnten, gesamteuropäischen Entwicklungen, mit hoher Wahrscheinlichkeit auch nicht vor dem österreichischen Kultursektor halt machen wird.

8.1. Veröffentlichung von *Linked Data*

Diskussion

Die Veröffentlichung der Metadaten des OBV hätte zur Folge, dass Verknüpfungen zu den von den österreichischen wissenschaftlichen Bibliotheken verwalteten Inhalten in das von der Semantic-Web Idee etablierte „Web of Data“ einfließen.

Die Vorteile dieser Vorgehensweise wären die erhöhte Sichtbarkeit der österreichischen Bibliotheksdaten im WWW, die auch einen erhöhten Besucherfluss auf die jeweiligen institutionellen Web-Auftritte mit sich bringen kann. Darüber hinaus würde diese Maßnahme mit der aktuellen europäischen Agenda hinsichtlich der Verfügbarkeit öffentlicher Daten übereinstimmen, dadurch positives Medienecho erzeugen sowie interessante Kooperationsmöglichkeiten mit Open-Cultural-Data-Initiativen wie beispielsweise OpenGLAM⁵⁶ (GLAM=Galleries, Libraries, Archives, Museums) eröffnen. Die mit externen Datenquellen integrierte Verfügbarkeit der bereitgestellten Daten könnte für WissenschaftlerInnen verschiedener Disziplinen neue Forschungsperspektiven eröffnen und gerade auch im Bereich der sich etablierenden digitalen Geisteswissenschaften (Digital Humanities⁵⁷) eine wichtige Rolle spielen. Darüber hinaus könnten im Bereich der Creative-Industries sowie des Tourismus neue Verwertungsmöglichkeiten geschaffen werden.

Dem gegenüber stehen Bedenken hinsichtlich des Wegfalls von Einnahmequellen sowie des Verlustes der Kontrolle über die hochqualitativen Daten. Bezüglich Metadaten als Einnahmequelle wurde bereits in Abschnitt 3 erwähnt, dass führende europäische Bibliotheken wie die Deutsche und Französische Nationalbibliothek damit rechnen, dass es in Zukunft keinen Markt für den Verkauf von bibliothekarischen Metadaten mehr geben wird²³ und eine Perspektive als zusätzliche Einnahmequelle damit wegfällt. Beide Institutionen haben dementsprechend damit begonnen, Teile ihrer Daten frei zur Verfügung zu stellen. Die Deutsche Nationalbibliothek hat beispielsweise auch ihr Geschäftsmodell⁵⁸ explizit dahingehend geändert, dass ab voraussichtlich Mitte 2015 alle bisher teilweise kostenpflichtig bereitgestellten Daten frei abgerufen werden können. Das zweite Argument gegen die Veröffentlichung der Daten richtet sich an den Verlust der Kontrolle über bereitgestellte Daten, wenn sie unter einer freien Lizenz oder gar einem Rechteverzicht veröffentlicht werden. In der Tat ist es so, dass sobald Inhalte einmal im WWW verfügbar sind, potentiell jeder Internet-NutzerIn darauf Zugriff hat und es praktisch unmöglich ist vollständig zu verfolgen, was im Anschluss daran

mit den Daten geschieht²⁵. Es ist schwer, in dieser Hinsicht Aussagen über die positiven oder negativen Folgen zu treffen, da langfristige Erfahrungswerte noch fehlen. Im Allgemeinen herrscht innerhalb der Linked Library Data Community jedoch der Konsens, dass die positiven Aspekte hier überwiegen. Im Zuge der Diskussion über die freie Veröffentlichung der Metadaten des Südwestdeutschen Bibliotheksverbundes konnte beispielsweise jede Teilnehmereinrichtung ein Veto einlegen, wovon jedoch nur wenige Einrichtungen Gebrauch gemacht haben⁵⁹.

Konzeptionelle und technische Umsetzung

Hinsichtlich der konzeptionellen und technischen Umsetzung der Veröffentlichung der Verbunddaten als Linked Open Data kann von den bereits gesammelten und gut dokumentierten Erfahrungen bei der Konvertierung von MAB in RDF profitiert werden. Die kürzlich erschienenen ersten Empfehlungen zur Verwendung von Ontologien durch die DINI AG KIM Titeldatengruppe (siehe Kapitel 5.3.) enthalten bereits viele praktische Richtlinien, sodass die Konvertierung lediglich eine Frage der technischen Umsetzung wäre.

Die in RDF abgebildeten Daten könnten über eine Kombination von den in Kapitel 4.1 vorgestellten und diskutierten Diensten bereitgestellt werden, wobei insbesondere ein Download-Angebot des Gesamtbestandes nur unter einer offenen Lizenz sinnvoll wäre, da ansonsten fraglich ist, ob potentielle Nutzer den Aufwand betreiben, die Daten mit Verwendungshinweisen und Restriktion zu versehen (siehe Kapitel 3, Europeana). Der konkrete Nutzen der verschiedenen Möglichkeiten der Veröffentlichung und der damit verbundene Ressourcenaufwand sollte jedenfalls in einer gesonderten Machbarkeitsstudie erhoben werden.

8.2. Nutzung von Linked Data

Die Nutzung von nach den Linked Data-Prinzipien veröffentlichten Daten kann, ebenso wie die Veröffentlichung, in unterschiedlichen Ausformungen erfolgen, die sich drastisch in ihrer Komplexität unterscheiden. Ein Beispiel einer solchen Nutzung ist die Anzeige von Informationen aus der semantischen Version der Online-Enzyklopädie Wikipedia, wie sie in den österreichischen Primo-basierten Katalogen schon seit Jahren im Einsatz ist⁶⁰. Die technische Umsetzung dieses Anwendungsfalles gründet jedoch auf recht klassischen Technologien wie relationalen Datenbanken. Gemäß diesem Vorbild wäre es denkbar, weitere Datensets als Quellen für Anzeige von Kontextinformationen in Suchinterfaces einzubinden. Ein solches Abholen

und Einblenden von Inhalten von Drittanbietern erlaubt es dem Benutzer über die lokal vorgehaltenen Informationen hinaus zu recherchieren, wobei gleichzeitig sicher gestellt wird, dass die Suchportale der Bibliotheken nicht verlassen werden.

Als mögliche Datenquellen bieten sich alle oben genannten bereits veröffentlichten Bestände (Kapitel 4), die Normdatei GND, die Wikipedia, Geonames⁶¹ und zahlreiche andere, teilweise sehr fachspezifische Sets an. Ein beispielhafter Ablauf einer Suche könnte wie folgt aussehen: Ein Benutzer startet eine thematische Recherche in einer Suchoberfläche und erhält eine Ergebnisliste mit relevanter Literatur. Er öffnet die Detailanzeige eines besonders interessanten Treffers und wird sogleich mit kontextualisierten Informationen zum Autor (aus der GND) und zum Thema (ebenfalls aus der GND) versorgt. Darüber hinaus kann der vollständige Zugriff auf die Relationen innerhalb der GND genutzt werden, das Themengebiet des aktuellen Treffers innerhalb der Ordnung der GND zu verorten: es werden über-, unter- und gleichgeordnete Themengebiete angezeigt.

Die mithilfe von über die GND hinausgehenden Informationen kontextualisierten Inhalte der Bibliothekskataloge können darüber hinaus auch neue Formen der Suche ermöglichen und beispielsweise auch die geographische Verortung von Werken und deren Inhalten sowohl auf digitalen Karten als auch für mobile Anwendungen miteinschließen – Orte sind in der GND derzeit noch nicht mit Geokoordinaten versehen, was aber durch eine Verknüpfung mit Einträgen aus der Geonames-Datenbank erreicht werden kann. Auch soziale Relationen zwischen Personen (sowohl zwischen Autoren als auch zwischen Personen, die in den Werken erwähnt werden) können neue Wege der Navigation durch die stetig wachsenden Bibliotheksbestände erschließen. Auf solche Weise durchsuchbare Kataloge betten die Suche nach Dokumenten und Büchern in das allgemeine Informationsbedürfnis des Nutzers ein.

9. Bewertung/Fazit

Die Veröffentlichung des gesamten österreichischen Verbundkataloges bzw. von Teilen davon nach einer der oben beschriebenen Methoden hätte eine Reihe von Vorteilen: Zunächst würde die Sichtbarkeit dieser Daten gesteigert. Da bereits jetzt eine Reihe deutscher Verbünde den Schritt zu einer vollständigen Freigabe ihrer Daten vollzogen haben, könnte ein Zö-

gern in dieser Frage potentielle Anwender davon abhalten, die Daten des OBV nachzunutzen und dadurch indirekt auf ihn zu verlinken. Wie oben erwähnt, wurden auch im Rahmen des Projektes culturegraph.org, eine Reihe von Services angedacht, die teilweise nur mit offenen Katalogdaten zu realisieren wären. Für die Veröffentlichung der Daten sprechen aber auch noch andere Gründe, die teilweise mit dem Auftrag der Bibliotheken als öffentlich finanzierte Informationsdienstleister zusammenhängen: Ein sehr früh formuliertes Ziel des Semantic Web, das auf den verbundenen Datenbeständen basieren soll, ist die nicht vorhersehbare Verwendung von Daten und so könnten die gut gepflegten Daten des OBV neue Forschungsfelder eröffnen: Buchforschung, Verlagsgeschichte, Erhebungen zur Bücherverbreitung oder Erwerbungspolitik bzw. Digital Humanities im Ganzen seien als Schlagwörter genannt. Gerade im Bereich solcher Forschungsvorhaben würde die freie Verfügbarkeit der Metadaten die wissenschaftliche Auseinandersetzung mit ihnen drastisch vereinfachen, besonders beispielsweise auch die einfache Reproduzierbarkeit publizierter Ergebnisse ermöglichen.

Im Lichte der Open Access-Bewegung ist es schlussendlich kaum zu argumentieren, dass Forscher ihre Ergebnisse möglichst barrierefrei anbieten, die entsprechenden Metadaten jedoch unter Verschluss gehalten werden sollen. Folgt man dieser Logik, kann man argumentieren, dass der öffentlich finanzierte Bibliothekssektor als Dienstleister für die Wissensgesellschaft dazu angehalten werden sollte, die nicht unter urheberrechtlichem Schutz stehenden Metadaten frei zur Verfügung zu stellen. Im Gegenzug zu einer Veröffentlichung der eigenen Daten bietet sich bei entsprechender Teilnahme der Vorteil, dass wiederum Daten anderer Institutionen dazu genutzt werden, die eigenen Inhalte zu kontextualisieren.

Mit der GND steht die wohl wichtigste Datenquelle bereits zur Verfügung. Sie verweist über zahlreiche Links zu anderen Datenbeständen und kann so als Einstiegspunkt ins Semantic Web gesehen werden. Es ist zu erwarten, dass culturegraph.org eine ähnliche Rolle einnehmen wird. Die oben skizzierte Nutzung wäre ein für die gesamte LD-Gemeinschaft wichtiger Meilenstein, da sich die meisten der Projekte im Bibliotheksbereich bisher auf die Veröffentlichung konzentrierten. Es sei darauf hingewiesen, dass eine solche Nachnutzung nur deshalb möglich ist, weil die veröffentlichenden Einrichtungen eine lizenzrechtlich freie Nutzung erlauben – für den Erfolg von Linked Data insgesamt ist ein möglichst niedrigschwelliger Zugang zu den Daten daher erfolgskritisch.

10. Empfehlungen

Während Linked Open Data-Prinzipien bereits in Form von Nutzung bestehender externer Datenquellen (z.B. Wikipedia-Tab in Primo) eingesetzt werden, besteht bezüglich Veröffentlichung der Verbunddaten noch Entscheidungsbedarf. Eine entsprechende Debatte wurde in Österreich noch nicht in ausreichender Intensität geführt, weswegen die Arbeitsgruppe folgende Empfehlungen ausspricht:

1. Das nächste Ziel der Arbeitsgruppe ist die Entscheidungsvorbereitung über die Veröffentlichung der Verbunddaten unter einer freien Lizenz. Dazu bedarf es weiterer, intensiver Diskussionen aller betroffenen Parteien, die durch die Arbeitsgruppe moderiert werden können.
2. Die hier präsentierten Ergebnisse, d.h. dieser Endbericht, werden nach entsprechender Bearbeitung durch die Veröffentlichung in einem geeigneten wissenschaftlichen Medium einer breiteren Öffentlichkeit zugänglich gemacht.

Patrick Danowski
Institute for Science and Technology Austria
E-Mail: patrick.danowski@ist.ac.at

Dipl.-Ing. Doron Goldfarb
Österreichische Nationalbibliothek
E-Mail: doron.goldfarb@onb.ac.at

Mag. Verena Schaffner, MSc
Die Österreichische Bibliothekenverbund und Service GmbH (OBVSG)
E-Mail: verena.schaffner@obvsg.at

Dr. Wolfram Seidler
Universitätsbibliothek Wien
E-Mail: wolfram.seidler@univie.ac.at

Alle Links wurden zuletzt am 9. September 2013 abgerufen.

- 1 <http://www.w3.org/standards/semanticweb/>
- 2 <http://www.w3.org/DesignIssues/LinkedData.html>
- 3 <http://de.dbpedia.org>
- 4 <http://datendienst.dnb.de/cgi-bin/mabit.pl?userID=opendata&pass=opendata&cmd=login>
- 5 Beispiel: UB Mannheim, <http://blog.bib.uni-mannheim.de/Aktuelles/?p=3172>
- 6 Für eine ausführliche Betrachtung zum Schutz von Metadaten siehe Kreutzer, Till (2011): Open Data- Freigabe von Daten aus Bibliothekskatalogen. Studie erstellt im Auftrag des hbz. Online unter: <http://www.hbz-nrw.de/dokumentencenter/veroeffentlichungen/open-data-leitfaden.pdf>
- 7 <http://www.w3.org/DesignIssues/LinkedData.html>
- 8 <http://creativecommons.org/licenses/>
- 9 Vgl. auch Danowski, Patrick (2012): Kontext Open Access. Creative Commons. In: Mitteilungen der VÖB, 65 (2), 2012, S.200–212. Online unter: <http://phaidra.univie.ac.at/o:175746>
- 10 <https://wiki.dnb.de/display/DINIAGKIM/LLD+Charter>
- 11 http://ec.europa.eu/information_society/policy/psi/rules/eu/index_en.htm
- 12 http://europa.eu/rapid/press-release_IP-11-1524_en.htm
- 13 <http://de.scribd.com/doc/95437065/Topic-Report-Open-Data-in-Cultural-Heritage-Institutions>
- 14 <http://ec.europa.eu/digital-agenda/>
- 15 <http://pro.europeana.eu/pro-blog/-/blogs/1457697>
- 16 <http://www.europeana.eu>
- 17 <http://www.europeana-newspapers.eu/>
- 18 <http://dm2e.eu/>
- 19 <http://pro.europeana.eu/web/guest/data-exchange-agreement>
- 20 <http://pro.europeana.eu/documents/900548/8a403108-7050-407e-bd00-141c20082afd>
- 21 http://www.bnl.public.lu/fr/actualites/communiqués/2011/10/CENL/open-data_cenl_2011.pdf
- 22 Beispiel: <http://de.slideshare.net/brocadedarkness/libris-linked-library-data>
- 23 <http://conference.ifla.org/past/ifla78/181-niggemann-en.pdf>
- 24 <http://conference.ifla.org/past/ifla78/181-illien-en.pdf>
- 25 http://www.w3.org/2013/04/odw/odw13_submission_57.pdf
- 26 „Licence Ouverte“ - <http://ddata.over-blog.com/xxxxxy/4/37/99/26/licence/Licence-Ouverte-Open-Licence-ENG.pdf>
- 27 <http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf>
- 28 <http://bibliontology.com/>

- 29 <http://dublincore.org/>
30 <http://www.foaf-project.org/>
31 <http://www.w3.org/2004/02/skos/>
32 <http://viaf.org/>
33 <http://id.loc.gov/authorities/subjects.html>
34 <http://www.hbz-nrw.de/dokumentencenter/produkte/lod/aktuell/>
35 <http://www.hbz-nrw.de/dokumentencenter/presse/pm/datenfreigabe>
36 <https://wiki1.hbz-nrw.de/display/SEM/Aktuelle+Open-Data-Exporte>
37 http://datahub.io/dataset/hbz_unioncatalog
38 <http://lod.b3kat.de/doc>
39 <http://lod.b3kat.de/download>
40 http://www.hebis.de/de/1ueber_uns/projekte/lod/lod_index.php
41 http://www.hebis.de/de/1ueber_uns/projekte/lod/lod_download.php
42 Alle in den RDF-Tripeln der GND verwendeten Relationen sind nunmehr abrufbar unter: GND Ontology – <http://d-nb.info/standards/elementset/gnd>
43 British Library, Metadata Services, Standards. „Data Model Meeting.“ <http://www.bl.uk/bibliographic/meeting.html>.
44 <http://dcevents.dublincore.org/BibData/fyo>
45 http://wiki.dublincore.org/index.php/Bibliographic_Metadata_Task_Group#
46 <http://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>
47 http://www.kim-forum.org/Subsites/kim/DE/Home/home_node.html
48 <https://wiki.dnb.de/pages/viewpage.action?pagelD=68060017>
49 <http://dcpapers.dublincore.org/pubs/article/view/916/912>
50 <http://viaf.org>
51 <http://dewey.info/>
52 <http://www.deutsche-digitale-bibliothek.de/>
53 <http://www.w3.org/community/schemabibex/>
54 <http://igelu.org/special-interests/lod>
55 <http://gov.opendata.at/site/>
56 <http://openglam.org>
57 <http://www.dig-hum.de>
58 http://www.dnb.de/DE/Service/DigitaleDienste/Geschaeftsmodell/geschaeftsmodell_node.html
59 http://swop.bsz-bw.de/volltexte/2013/1067/pdf/bsz_newsletter_2013_1.pdf
60 Vgl. Knitel, Markus/Krabo, Ulrike (2011): Library Linked Data. Technologien, Projekte, Potentiale. In: Mitteilungen der VÖB, 64 (1), 2011, S.11–31. Online unter: http://www.obvsg.at/fileadmin/files/obvsg/publ/krabo_knitel_linked_data2.pdf
61 <http://www.geonames.org>